Free-floating bike share as a last mile transit connection: using hazard models to understand bike share patterns at UBC

M. Jake Vissers
August, 2020
School of Community and Regional Planning, UBC
Graduate Capstone Project

SCARP
School of Community
And Regional Planning

UBC

TABLE OF CONTENTS

TABLE OF FIGURES

ABSTRACT

The University of British Columbia (UBC) launched its first permanent free-floating bike share (FFBS) in 2019 with the arrival of HOPR. This new system is part of a global trend which has seen FFBS grow significantly in the last five years. As such, exploring how this emerging technology is being used has become more critical in planning for existing and future systems. Existing literature shows that proximity to transit can increase demand for FFBS, suggesting that FFBS may be an effective first/last mile connection. The purpose of this study is to determine whether proximity to frequent transit increases FFBS demand at UBC. To do this, I examine how ambient factors, such as weather, land use and transit, among others, affect bicycle idle duration, which I use as an inverse proxy for demand. This is done by estimating a Cox proportional hazard model, which models how each covariate affects the probability of a bicycle being booked. This study offers a unique exploration of how FFBS is used on a major university campus, particularly whether the system serves as a last mile connection for transit. These findings may benefit future planning decisions at UBC, as well as other universities interested in FFBS.

I. INTRODUCTION

Bike share systems have grown significantly in the past decade, becoming an increasingly important piece of urban transportation networks. This growth has been further fueled in the last five years by the development of a new bike share technology: free-floating bike share (FFBS). While traditional docked bike shares require stations for both locking and payment, FFBS integrates both critical features into the bicycle itself, allowing for more flexibility and lower operating costs. These advantages have helped spur considerable growth worldwide. This is perhaps most notable in China, where the two largest FFBS companies, OFO and Mobike, have already collectively deployed roughly 15 million bikes and serve over 300 million users [6].

Understanding use patterns for this new technology is critical for cities, universities and operators to best utilize them. Because demand is often asymmetric within a system, bikes in a FFBS are often rebalanced to increase utilization, particularly around major transit stations during peak demand [8]. Additionally, a better understanding of FFBS use patterns could aid cities and universities in planning for bicycle infrastructure, transit and public realm management.

Existing studies have posed a breadth of questions on the subject, examining both docked and free-floating bike shares. A 2017 study examined the distribution of bikes in a FFBS throughout the day to optimize the redistribution of bikes to meet demand [10]. The effects of weather, time of day, population density, land use and transportation infrastructure on use patterns have also been explored in Singapore [12] and Shanghai [6], among other cities [3][14]. Furthermore, evidence exists that bike sharing systems may enhance transit usage [7][16], and conversely, the existence of transit can benefit bike share. Specifically, recent studies have found that the presence of transit has a positive effect on trip attractions [14] and frequency of use. These results suggest that bike share may serve as a first/last mile connection with transit [5].

This paper investigates FFBS at the University of British Columbia (UBC) to determine whether proximity to transit increases demand. The study examines how ambient factors, such as weather, land use and transit, among others, affect bicycle idle duration, which is used as an inverse proxy for demand. Four months of FFBS trip data from August-November 2019 are utilized to estimate a Cox proportional hazard model. This model estimates the effect each covariate has on the probability of a bicycle being booked. Assuming that a higher booking probability indicates more demand, the model helps unpack how transit affect FFBS demand, and provides clues to whether bike share is being used as a last mile transit connection on campus.

Each HOPR bicycle generates GPS trip data that is linked to a unique bicycle ID, providing detailed trips of each individual bicycle. This data was provided by Campus and Community Planning, as laid out by the HOPR licensing agreement. Transit, land use and weather data was also publicly available to use in this study.

II. DATA AND METHODS

A. Data

The analysis uses data from HOPR, a FFBS at UBC. The original dataset contained information on 29,957 trips between August and November 2019. This data was used to generate a bicycle idle duration dataset, which tracks the duration that each bicycle was not in use. The variable idle_duration serves as the dependent variable in this analysis, which assumes an inverse relationship between idle duration and bike share demand.

To clean the data, I decided to only include points in which a bicycle stays in one place, within the campus service area during the entire time it remained idle. This removed any points where the bicycle was moved by the operator for maintenance or rebalancing, or was tampered with between trips. Once removed, the final dataset contained 23,363 points.

A full set of variables can be found in Table I.

B. Descriptive Statistics

The variation in bicycle idle duration is illustrated in Figure 1 and Figure 2. Figure 1 demonstrates the distribution of idle durations, and shows that the majority of bikes idle for no more than 6 hours. Figure 2 shows the same distribution of idle durations with a cumulative distribution. The vertical red line marks 72 hours, highlighting that nearly 100 percent of bikes are checked out within 3 days.

Figure 3 explores the geospatial distribution of bicycle idle durations by showing the mean idle duration per 200 meter zones across campus. There are clearly a few patterns shown in this map. The academic core of campus appears to have shorter idle durations. Additionally, pockets around major transit hubs on the east side of campus show shorter idle durations, but the pattern is less clear. Finally, large pockets in southeast campus around the Wesbrook Village and Hampton Place neighborhoods exhibit higher idle durations, indicating lower system demand.

Figure 4 examines the geospatial distribution of idle durations across 4 temporal frames: all idle durations, those longer than 24 hours, those longer than 48 hours and those longer than 72 hours. While only 9.3% of bicycles idle longer than 24 hours, it's evident from Figure 4 that the spatial patterns are different for bikes with long idle durations. For example, Figure 4.d. shows a cluster of bikes in the southeast corner of campus, indicating that longer idle durations disproportionately occur there.

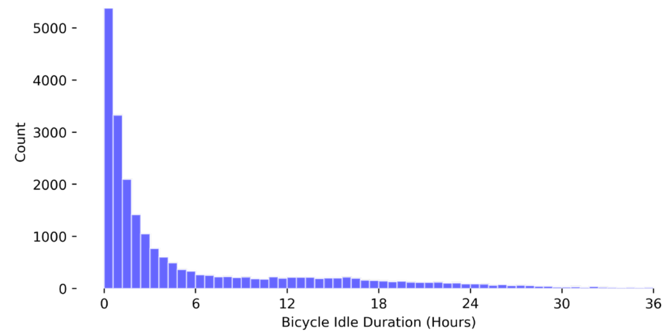


Figure 1 - Histogram of idle duration (hours) for individual bikes

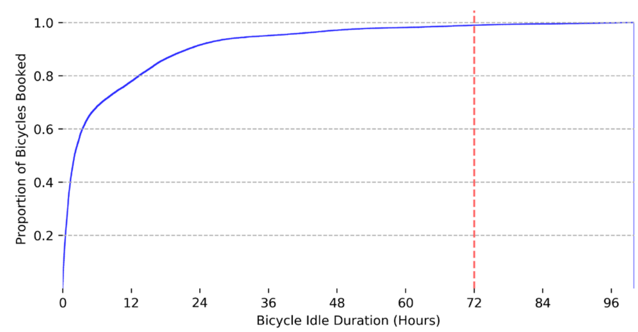


Figure 2 - Cumulative distribution of idle duration (hours)

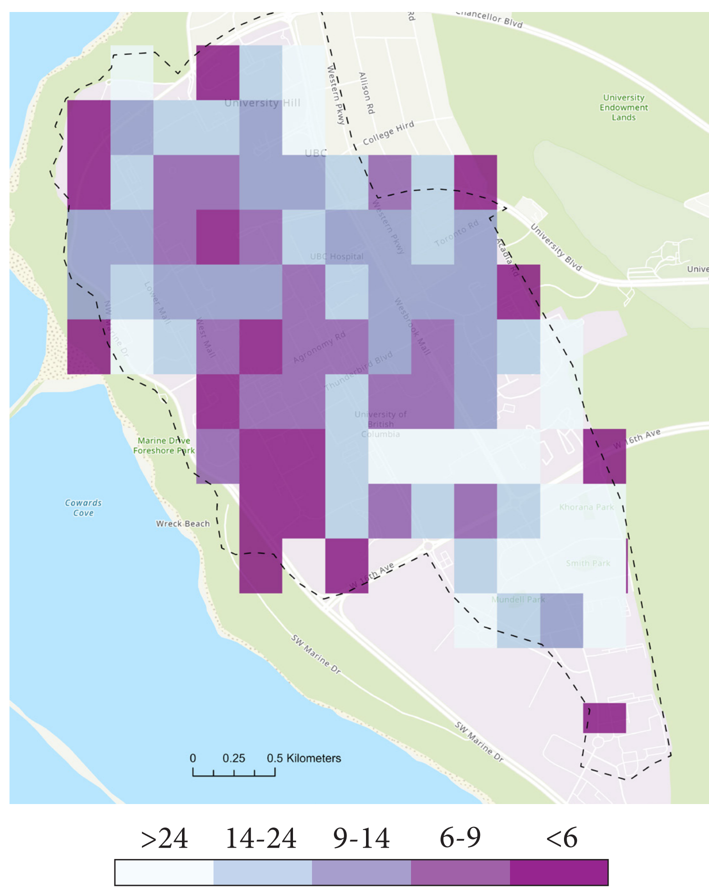
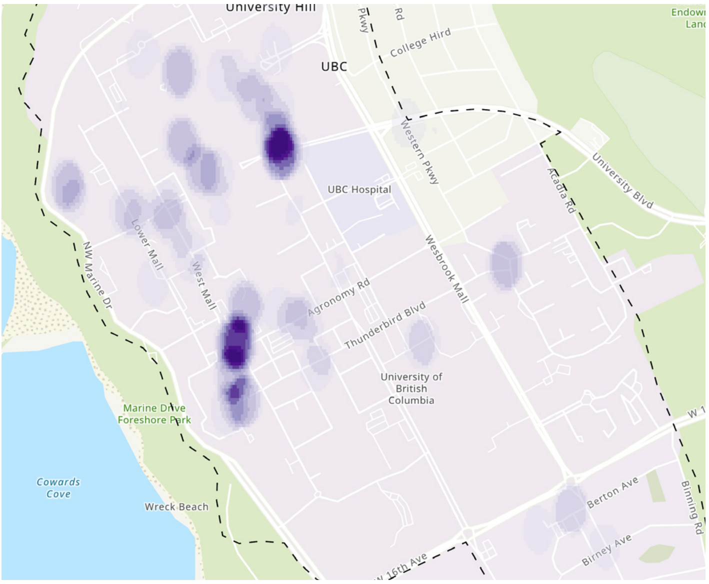


Figure 3 - Mean idle duration (hours) per 200 meter zones



a. All idle durations (N = 23363)

b. Idle durations > 24 hours (N = 2176)

c. Idle durations > 48 hours (N = 884)

d. Idle durations > 72 hours (N = 452)

Figure 4 - Idle duration heatmaps for four temporal frames

C. Cox Proportional Hazard Model

A Cox proportional hazard model was used to explore the effect of several spatial, temporal and weather variables on the idle duration of bicycles. The model estimates the probability of an event occurring at a given time. For this model, the event will be a bicycle being booked, ending the idle duration of that bike. Therefore, the model can be said to give the probability of an individual bicycle being booked at a given time.

The Cox proportional hazard model is built around the proportional hazard assumption, which says that all individuals have the same baseline hazard function, scaled with a unique factor. Violations of this assumption can be tested with Schoenfeld residuals, along with a combination of statistical and visual tests.

The general form for a Cox proportional hazard model is as follows:

$$h(t|x) = h_0(t)\, exp\left(\sum_{i=1}^{n} \beta_i x_i\right)$$

The covariates (x_i) are assumed to have a proportional effect on a population-level baseline hazard ($h_0(t)$) to generate a time-dependent individual hazard function. Additionally, the likelihood of an individual bicycle being checked out at a time i can be written as follows:

Description	Variable Name	Min	1st Quartile	Median	Mean	3rd Quartile	Max
Idle duration of a bicycle (hours)	idle_duration	0.0	0.7	2.1	9.4	10.6	633.5
Proportion of idle duration within 50 meters of another bikes	nearby_time	0.0	0.0	0.4	0.9	1.3	12.4
Mean temperate during idle duration (˚C)	avg_temp	-2.5	8.5	11.1	12.0	15.8	27.0
Frequency of transit around each bike (stops per hour ÷ distance² (m²))	transit_index	0.0	0.1	0.1	6.6	0.3	143666.4
Total building density around each bike (building footprint (m²)÷ distance² (m²))	building_index	0.3	1.7	2.2	2.6	2.7	1604.8
Building index for residential buildings	residential	0.1	0.2	0.3	0.8	0.7	1604.2
Building index for academic buildings	academic	0.2	0.2	0.5	0.8	1.1	77.0
Building index for parkades	parking	0.0	0.1	0.1	0.2	0.2	59.5
Building index for other buildings	other	0.0	0.2	0.2	0.4	0.4	318.2
Precipitation dummy	precipitation	22.4% of bookings					
Peak time dummy (booking occurs M-F, 8am-6pm)	peak_time	47.7% of bookings					

Table I - Variable summary

$$L_i(\beta) = \frac{exp(\beta x_i)}{\sum_{j:t_j>t_i} exp(\beta x_j)}$$

Further, the partial likelihood is found through the product of individual likelihood functions across all events. The log of the following function can be maximized to estimate model parameters [2].

$$L(\beta) = \prod_{i=1}^{D} L_i(\beta)$$

D. Model Specification

An initial model was specified which included the following variables from Table I: nearby_time, peak_time, avg_temp, precipitation, transit_index and building_index. However, the variables nearby_time, peak_time, avg_temp and precipitation failed a statistical test for the proportional hazard assumption, and would require further investigation.

To explore the nearby_time variable, I split the data into quartiles and fit each one to a Nelson-Aalen cumulative hazard curve (Figure 5). The Nelson-Aalen fitter uses the following function to estimate the cumulative hazard rate, where 'd_i' represents events occurred and 'r_i' represents the total population at risk [1].

$$\widehat{H}(t) = \sum_{t_i \le t} \frac{d_i}{r_i}$$

These four curves shows the variation in hazard across the four quartile of the nearby_time variable. From this plot it's clear that bikes with the most time spent near other bikes (nearby_time>1.3) have the lowest cumulative hazard rate. Conversely, bikes with no time spent near other bikes (nearby_time = 0), have the highest cumulative hazard rate, confirming the negative relationship between nearby_time and hazard found in Table II. However, the uneven change in cumulative hazard between each quartile suggests a non-linear relationship may exist. As such, a nearby_time^2 term was added to the final model.

The dummy variables peak_time and precipitation were changed to fixed variables to address the proportional hazard assumption. This would allow the covariates to be included in the model without estimating their effect.

The variable avg_temp failed the proportional hazard test as well. However, the Schoenfeld residual plot (Figure 6) showed an even distribution of residuals, indicating that any assumption violations were likely minor. It was determined that the covariate is acceptable to remain in the model as is [13].

Additionally, the covariate building_index was divided into four categories, residential, academic, park-

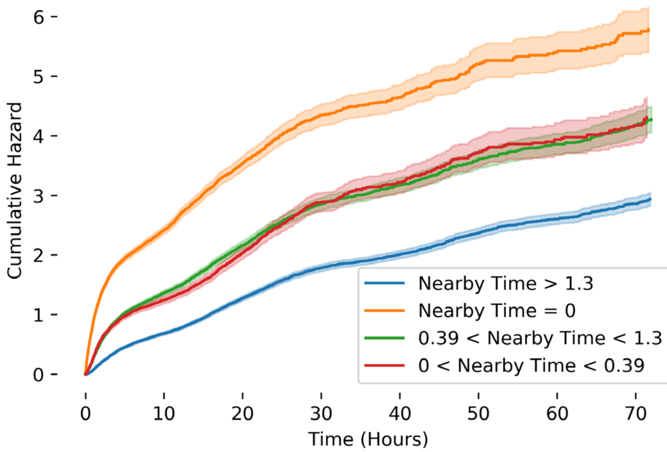


Figure 5 - Nelson-Aalen cumulative hazard estimator for nearby time

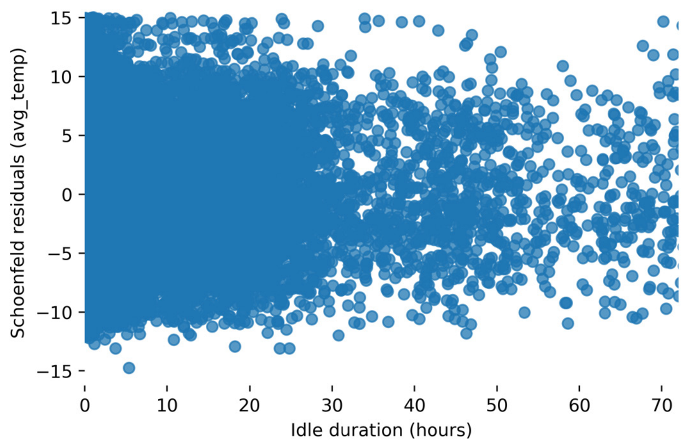


Figure 6 - Schoenfeld residuals for avg temp

ing, and other, to investigate the impact of various building types. The functional form for this model can be found with Equation 1.

$$h_i(t) = h_0(t_j)\, exp\big(\beta_1 nearby\_time_i + \beta_2 nearby\_time_i{}^2 + \beta_3 avg\_temp_i + \beta_4 transit\_index_i + \beta_5 residential_i + \beta_6 academic_i + \beta_7 parking_i + \beta_8 other_i\big)$$

Equation 1 - Model 1

A second model was also estimated with the transit_index variable broken up into four quartiles and included as dummy variables relative to the first quartile. Initially, the new transit specification returned unexpected results that were inconsistent with literature. However, it turned out that unaccounted variables were influencing the model.

Mapping high transit_index across the study area (Figure 7) highlighted a few patterns that could potentially impact model results. First, it was clear that high transit_index often occurred along high traffic and volume roadways, which may be more hostile to cyclists than other parts of campus which are pedestrianized or traffic calmed. Second, a high transit_index pocket was present at the core of Wesbrook Village, a university neighborhood located at the south end of campus. Although the neighborhood has a high transit_index, it is a primarily residential and commercial development over a kilometer south of the campus's academic core. Because of this, it is reasonable to suspect that use patterns could differ from main campus.

To account for both of these potential effects, dummy variables were created for idle bicycles in Wesbrook Village and within 50 meters of a 50 km/h or higher road. These variables are included in Model 2 (Equation 2), and visualized in Figure 9.

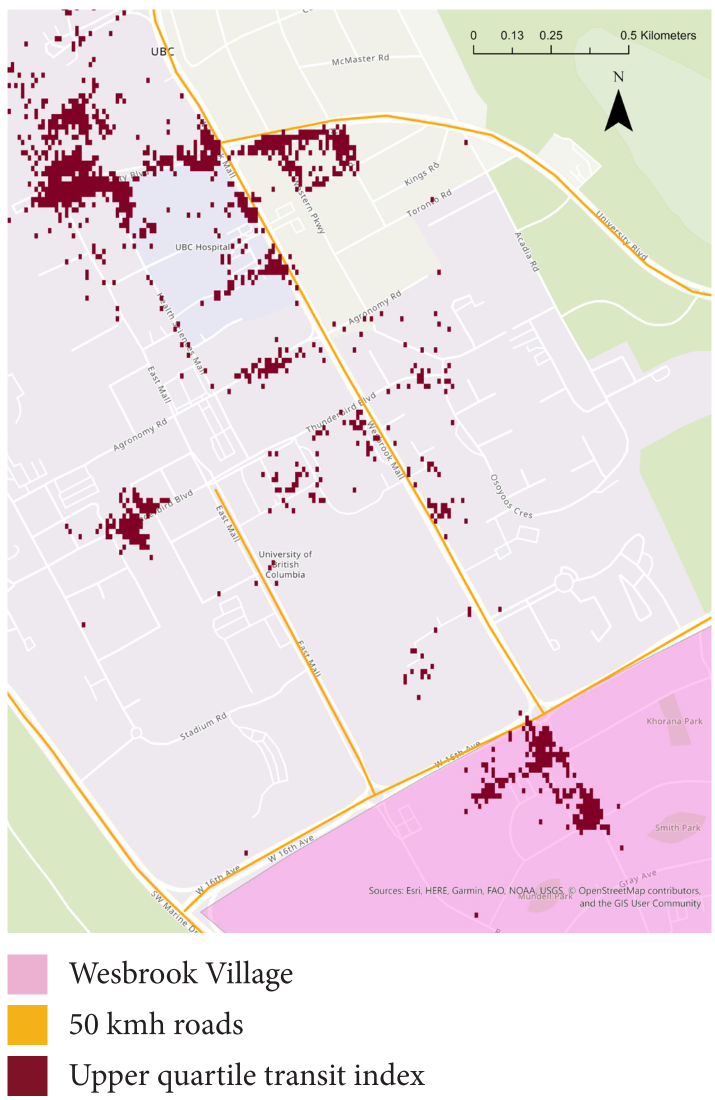


Figure 7 - Upper quartile transit index

	Model 1				Model 2			
	Coef		SE	HR	Coef		SE	HR
nearby_time	-0.45	***	0.01	0.63	-0.46	***	0.01	0.63
nearby_time^2	0.05	***	0.00	1.05	0.05	***	0.00	1.05
avg_temp	0.02	***	0.00	1.02	0.02	***	0.00	1.02
transit_index	0.00		0.00	1.00	-		-	-
transit_2	-		-	-	0.15	***	0.02	1.16
transit_3	-		-	-	0.13	***	0.02	1.14
transit_4	-		-	-	0.20	***	0.02	1.22
residential	0.00		0.00	1.00	0.00		0.00	1.00
academic	0.02	***	0.00	1.02	0.01	***	0.00	1.01
parking	-0.47	***	0.04	0.62	-0.68	***	0.05	0.51
other	0.00		0.00	1.00	0.00		0.01	1.00
busy_road	-		-	-	-0.45	***	0.03	0.64
wesbrook	-		-	-	-0.43	***	0.03	0.65
N	23363				23363			
Concordance	0.674				0.672			
Partial AIC	361768				361343			
Partial Log-Likelihood	-180876				-180659			
Pseudo R^2	0.145				0.146			
Significance Codes: p<0.1 *, p<0.05 **, p <0.01 ***								

Table II- Model summary

III. RESULTS

A Cox proportional hazard model was estimated for each of the specifications described in Equation 1 and Equation 2. The full results for these estimations can be found in Table II, and can be understood as the probability of a bicycle being booked at a given time. A McFadden pseudo r^2 value of 0.146 was calculated for Model 2, relative to the null log likelihood. This value was higher than that returned in Model 1, suggesting a better fit for the data. Further, a correlation matrix (Figure 8) showed no significant variable interaction for Model 2.

Time spent near other bicycles has a significant negative effect on booking probability, although the positive coefficient for the quadratic term nearby_time^2 diminishes the effect as the term nearby_time increases. Average temperature and proximity to academic buildings both have significant positive effects on booking probability, while proximity to parkades has a significant negative effect. However, proximity to residential buildings and other building types showed no significant effect.

Relative to quartile 1, both quartile 2 and 3 of transit_index returned a positive effect, but their respective coefficients were not significantly different from each other. Quartile 4, however, returned the highest significant positive effect on booking probability, indicating proximity to highest frequency transit most increases bike share demand. Additionally, proximity to busy roads and Wesbrook Village each had a significant negative effect on booking probability.

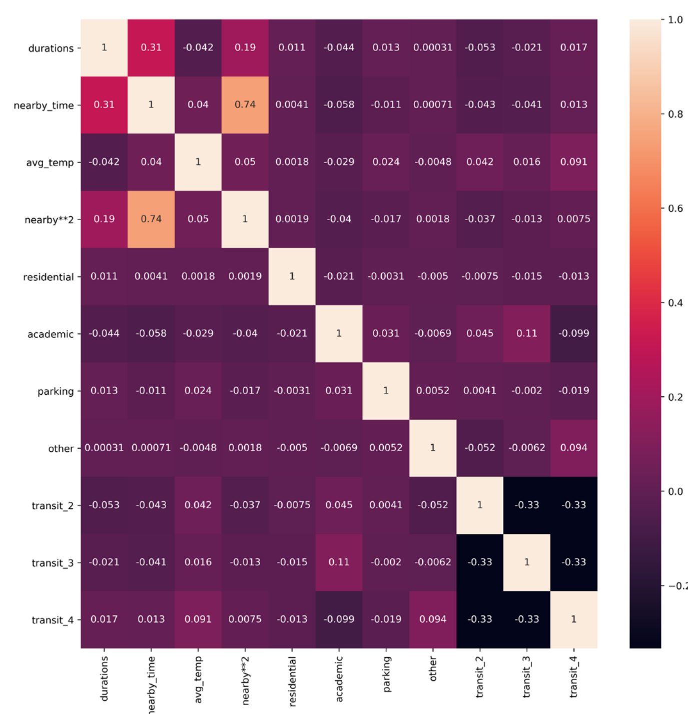


Figure 8 - Variable correlation matrix

Wesbrook Village

Busy road 50 meter buffer

Figure 9 - Wesbrook and busy road dummy variables

$$
\begin{aligned}
h_i(t) = h_0(t_j)\, exp\big(&\beta_1 nearby\_time_i + \beta_2 nearby\_time_i^{\,2} \\
&+ \beta_3 avg\_temp_i + \beta_4 transit\_2_i \\
&+ \beta_5 transit\_3_i + \beta_6 transit\_4_i \\
&+ \beta_7 residential_i + \beta_8 academic_i \\
&+ \beta_9 parking_i + \beta_{10} other_i + \beta_{11} busy\_road_i \\
&+ \beta_{12} wesbrook_i\big)
\end{aligned}
$$

Equation 2 - Model 2

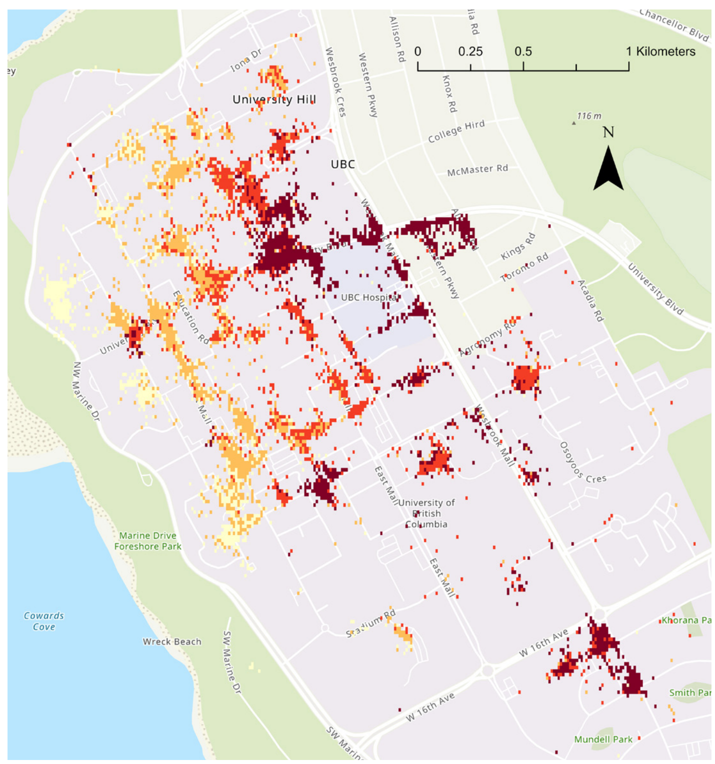
IV. DISCUSSION

Residential buildings had no effect on booking probability, which was an unexpected result. While much literature links population and building density to increased bike share use [3][5], the context at UBC is quite different to a city. Many residential buildings, particularly undergraduate residences, are close enough to the main academic core that walking may be preferable.

Wesbrook Village, a mixed-use neighborhood south of campus, had a significant negative effect on booking probability. This could be for a few reasons. Although Wesbrook Village has transit, the neighborhood may be too small for first and last mile bicycle connections. Further, the neighborhood is over a kilometer south of the campus core, and can only be reached by traveling on high speed roads (see Figure 9). It is possible that facilitating more bike share trips in Wesbrook Village would require more robust bicycle infrastructure between the neighborhood and main campus.

Another interesting finding was the significant negative effect of parkades on booking probability. This finding indicates that bike share is likely not being used as a last mile connector for automobile trips to campus.

The central question of this research, however, was to determine the effect of transit on bicycle booking probability. Assuming higher booking probability indicated more bike share demand, it's clear that closer proximity to high levels of transit has a positive effect on bike share demand and may be an effective last mile



1st Quartile

2nd Quartile

3rd Quartile

4th Quartile

Figure 10 - Four quartiles of transit index

connection at UBC. This result concurs with existing literature which shows a positive relationship between bike share and transit [3][5][7][14].

However, while the highest quartile of the transit_index variable saw the highest effect, the second and third quartile saw positive effects that were not significantly different. Mapping all four quartiles (Figure 10) clarified that the second and third quartiles of transit_index appeared at places close to high-frequency transit, but not immediately adjacent. This result likely indicates that bike share is most effective as a last mile transit connection when it is available directly beside transit stops. Additionally, the negative effect of high speed roads indicates frequent transit on traffic calmed roads may further improve bike share as a last mile connection.

By assessing the effect of transit, along with other factors, on bike share demand, his study improves the understanding of the HOPR network and could further advise future planning decisions. Given the unique location and geography of UBC, this study could also help inform FFBS planning in greater Vancouver, as well as other universities.

V. LIMITATIONS

The study has a few potential limitations that could have impacted the results. The bike share system has a transit exclusion zone around one bus loop, preventing bikes from parking there, and also has a parking hub system, which encourages users to park bicycles in designated hubs. Although these factors may bias where bicycles are parked, they should not impact the relationship between ambient factors and idle duration.

In addition to these system conditions, more detailed data could improve the predictive capability of the model. For example, adding a covariate for bicycle infrastructure could be more precise than a busy_road dummy variable, and further explain the negative effect of bicycles at Wesbrook Village. Additionally, rather than frequency, investigating ridership counts could be a better measure for the transit_index variable.

VI. ACKNOWLEDGMENTS

I would like to thank Campus & Community Planning and HOPR for providing access to data and answering questions throughout the study.

References

1. Aalen, O.O. (1978). Nonparametric inference for a family of counting processes, Annals of Statistics 6, 701–726
2. Cox, D.R., (1972). “Regression Models and Life-Tables”. Journal of the Royal Statistical Society, Series B. 34 (2), 187–220.
3. Faghih-Imani, A., Eluru, N., El-Geneidy, A. M., Rabbat, M., & Haq, U. (2014). How land-use and urban form impact bicycle flows: Evidence from the bicycle-sharing system (BIXI) in Montreal. Journal of Transport Geography, 41, 306–314. https://doi.org/10.1016/j.jtrangeo.2014.01.013
4. Guidon, S., Becker, H., & Axhausen, K. (2019). Avoiding stranded bicycles in free-floating bicycle-sharing systems: Using survival analysis to derive operational rules for rebalancing. 2019 IEEE Intelligent Transportation Systems Conference (ITSC), 1703–1708. https://doi.org/10.1109/ITSC.2019.8916869
5. Li, X., Du, M., & Yang, J. (2020). Factors Influencing the Access Duration of Free-floating Bike Sharing as a Feeder Mode to the Metro in Shenzhen. Journal of Cleaner Production, 123273. https://doi.org/10.1016/j.jclepro.2020.123273
6. Li, Y., & Shuai, B. (2018). Origin and destination forecasting on dockless shared bicycle in a hybrid deep-learning algorithms. Multimedia Tools and Applications, 79(7), 5269–5280. https://doi.org/10.1007/s11042-018-6374-x
7. Martens, K. (2007). Promoting bike-and-ride: The Dutch experience. Transportation Research Part A: Policy and Practice, 41(4), 326–338. https://doi.org/10.1016/j.tra.2006.09.010
8. Médard de Chardon, C., Caruso, G., & Thomas, I. (2016). Bike-share rebalancing strategies, patterns, and purpose. Journal of Transport Geography, 55, 22–39. https://doi.org/10.1016/j.jtrangeo.2016.07.003
9. NelsonAalenEstimator.pdf. (n.d.). Retrieved August 8, 2020, from http://www.medicine.mcgill.ca/epidemiology/hanley/c609/Material/NelsonAalenEstimator.pdf
10. Pal, A., & Zhang, Y. (2017). Free-floating bike sharing: Solving real-life large-scale static rebalancing problems. Transportation Research Part C: Emerging Technologies, 80, 92–116. https://doi.org/10.1016/j.trc.2017.03.016
11. Saberi, M., Ghamami, M., Gu, Y., Shojaei, M. H. (Sam), & Fishman, E. (2018). Understanding the impacts of a public transit disruption on bicycle sharing mobility patterns: A case of Tube strike in London. Journal of Transport Geography, 66, 154–166. https://doi.org/10.1016/j.jtrangeo.2017.11.018
12. Shen, Y., Zhang, X., & Zhao, J. (2018). Understanding the usage of dockless bike sharing in Singapore. International Journal of Sustainable Transportation, 12(9), 686–700. https://doi.org/10.1080/15568318.2018.1429696
13. Stensrud, M. J., & Hernán, M. A. (2020). Why Test for Proportional Hazards? JAMA, 323(14), 1401–1402. https://doi.org/10.1001/jama.2020.1267
14. Wang, K., & Chen, Y.-J. (2020). Joint analysis of the impacts of built environment on bikeshare station capacity and trip attractions. Journal of Transport Geography, 82, 102603. https://doi.org/10.1016/j.jtrangeo.2019.102603
15. Zamir, K. R., Bondarenko, I., Nasri, A., Brodie, S., & Lucas, K. (n.d.). Comparative Analysis of User Behavior of Dock-Based vs. Dockless Bikeshare and Scootershare in Washington, D.C. 23.
16. Zhang, Y., & Zhang, Y. (2018). Associations between Public Transit Usage and Bikesharing Behaviors in The United States. Sustainability, 10(6), 1868. https://doi.org/10.3390/su10061868